

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

JC921 U.S. PTO
09/748542
12/26/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application:

1999年12月27日

出願番号
Application Number:

平成11年特許願第370413号

出願人
Applicant(s):

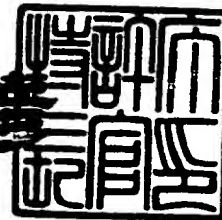
インターナショナル・ビジネス・マシーンズ・コーポレイシ
ョン

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 3月10日

特許庁長官
Commissioner,
Patent Office

近藤隆彦



出証番号 出証特2000-3014470

【書類名】 特許願

【整理番号】 JA999251

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/00

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

【氏名】 伊東 信泰

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

【氏名】 西村 雅史

【特許出願人】

【識別番号】 390009531

【住所又は居所】 アメリカ合衆国 1 0 5 0 4、ニューヨーク州アーモンク
(番地なし)

【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

【選任した代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【手数料の表示】

【予納台帳番号】 024154

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識装置、方法、コンピュータ・システム及び記憶媒体

【特許請求の範囲】

【請求項 1】

アナログ音声入力信号をデジタル信号に変換処理を行う音響処理手段と、
音の特徴を学習した音響モデルを記憶した記憶手段と、

予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第 1 の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第 2 の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶手段と、

前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する手段と、

を有する音声認識装置。

【請求項 2】

前記第 1 及び第 2 の言語モデルは N - g r a m モデルである請求項 1 記載の音声認識装置。

【請求項 3】

アナログ音声の入力手段と、

前記アナログ音声をデジタル信号に変換処理する手段と、

音の特徴を学習した音響モデルを記憶した記憶手段と、

予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第 1 の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第 2 の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶手段と、

前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する手段と、

前記認識された結果を表示する表示装置と

を有するコンピュータ・システム。

【請求項 4】

前記第 1 及び第 2 の言語モデルは N - g r a m モデルである請求項 3 記載のコンピュータ・システム。

【請求項 5】

アナログ音声入力信号をデジタル信号に変換処理し、

予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第 1 の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第 2 の言語モデルと、の両方の言語モデルを有する辞書を記憶し、

前記デジタル信号について音響モデルと前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する音声認識方法。

【請求項 6】

前記第 1 及び第 2 の言語モデルは N - g r a m モデルである請求項 5 記載の音声認識方法。

【請求項 7】

アナログ音声を入力し、

前記アナログ音声をデジタル信号に変換処理し、

予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第 1 の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第 2 の言語モデルと、の両方の言語モデルを有する辞書を記憶し、

前記デジタル信号について音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識し、

前記認識された結果を表示する音声認識方法。

【請求項 8】

前記第 1 及び第 2 の言語モデルは N - g r a m モデルである請求項 7 記載の音声認識方法。

【請求項 9】

コンピュータ・プログラムを有するコンピュータ読みとり可能な記憶媒体であって、前記記憶媒体は、

音響モデルと、

冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習し

た第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書

を記憶し、さらに前記コンピュータ・プログラムは、コンピュータに入力されたアナログ音声入力信号が変換されたデジタル信号について前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させるものである記憶媒体。

【請求項10】

前記第1及び第2の言語モデルはN-gramモデルである請求項9記載の記憶媒体。

【請求項11】

冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶媒体。

【請求項12】

前記第1及び第2の言語モデルはN-gramモデルである請求項11記載の記憶媒体。

【請求項13】

冗長語と冗長語以外の通常の単語（通常単語）とを含む文章について音声認識をするための装置であって、

- (1) 認識の対象となる入力された語が通常単語かどうか判断する手段と、
 - (2) 前記(1)で通常単語であると判断された場合、さらに入力された語を認識するために必要な条件となる語が通常単語のみからなるかどうかを判断する手段と、
 - (3) 前記(2)で条件となる語が通常単語のみでなく冗長語も含むと判断された場合、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させる手段と、
- を有する音声認識装置。

【請求項 14】

前記装置はさらに、

(1-1) 前記(1)で、対象となる語が通常単語でないと判断された場合、第1の言語モデルに基づいて確率を計算する手段を有する請求項13記載の音声認識装置。

【請求項 15】

前記装置はさらに、

(2-1) 前記(2)で条件となる語が通常単語のみからなると判断された場合、第2の言語モデルに基づいて確率を計算する手段を有する請求項13又は14記載の音声認識装置。

【請求項 16】

前記(3)はさらに、

(3-1) 対象となる語の直前が冗長語かどうかを判断する手段と、

(3-2) 前記(3-1)で、直前が冗長語であると判断された場合、第1の言語モデルと第2の言語モデルから確率を計算する手段と、
を有する請求項13ないし15のいずれか記載の音声認識装置。

【請求項 17】

前記装置はさらに、

(3-1-2) 前記(3-1)で、直前の語が冗長語でないと判断された場合、第2のモデルから確率を計算する手段を有する請求項13ないし16のいずれか記載の音声認識装置。

【請求項 18】

冗長語と冗長語以外の通常の単語(通常単語)とを含む文章について音声認識をするための方法であって、

(1) 認識の対象となる入力された語が通常単語かどうか判断するステップと、

(2) 前記ステップ(1)で通常単語であると判断された場合、さらに入力された語を認識するために必要な条件となる語が通常単語のみからなるかどうかを判断するステップと、

(3) 前記ステップ(2)で条件となる語が通常単語のみでなく冗長語も含むと

判断された場合、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させるステップと、

を有する音声認識方法。

【請求項19】

前記方法はさらに、

(1-1) 前記ステップ(1)で、対象となる語が通常単語でないと判断された場合、第1の言語モデルに基づいて確率を計算するステップを有する請求項18記載の音声認識方法。

【請求項20】

前記方法はさらに、

(2-1) 前記ステップ(2)で条件となる語が通常単語のみからなると判断された場合、第2の言語モデルに基づいて確率を計算するステップを有する請求項18または19記載の音声認識方法。

【請求項21】

前記ステップ(3)はさらに、

(3-1) 対象となる語の直前が冗長語かどうかを判断するステップと、
(3-2) 前記ステップ(3-1)で、直前が冗長語であると判断された場合、第1の言語モデルと第2の言語モデルから確率を計算するステップと、
を有する請求項18ないし20のいずれか記載の音声認識方法。

【請求項22】

前記方法はさらに、

(3-1-2) 前記ステップ(3-1)で、直前の語が冗長語でないと判断された場合、第2のモデルから確率を計算するステップを有する請求項18ないし21のいずれか記載の音声認識方法。

【請求項23】

冗長語と冗長語以外の通常の単語（通常単語）とを含む文章について音声認識

をするためのコンピュータ・プログラムを有するコンピュータ読みとり可能な記憶媒体であって、前記コンピュータ・プログラムは、コンピュータに

- (1) 認識の対象となる入力された語が通常単語かどうか判断する手順と、
 - (2) 前記手順(1)で通常単語であると判断された場合、さらに入力された語を認識するために必要な条件となる語が通常単語のみからなるかどうかを判断する手順と、
 - (3) 前記手順(2)で条件となる語が通常単語のみでなく冗長語も含むと判断された場合、冗長語と冗長語以外の通常の単語(通常単語)とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させる手順と、
- を実行させるものである記憶媒体。

【請求項 24】

前記プログラムはさらにコンピュータに、

- (1-1) 前記手順(1)で、対象となる語が通常単語でないと判断された場合、第1の言語モデルに基づいて確率を計算する手順を実行させるものである請求項 23 記載の記憶媒体。

【請求項 25】

前記プログラムはさらにコンピュータに、

- (2-1) 前記手順(2)で条件となる語が通常単語のみからなると判断された場合、第2の言語モデルに基づいて確率を計算する手順を実行させるものである請求項 23 または 24 記載の記憶媒体。

【請求項 26】

前記プログラムはさらにコンピュータに、

- (3-1) 対象となる語の直前が冗長語かどうかを判断する手順と、
 - (3-2) 前記手順(3-1)で、直前が冗長語であると判断された場合、第1の言語モデルと第2の言語モデルから確率を計算する手順と、
- を実行させるものである請求項 23 ないし 25 のいずれか記載の記憶媒体。

【請求項 27】

前記プログラムはさらにコンピュータに、
 (3-1-2) 前記手順(3-1)で、直前の語が冗長語でないと判断された場合、第2のモデルから確率を計算する手順を実行させるものである請求項23ないし26のいずれか記載の記憶媒体。

【請求項28】

アナログ音声入力信号をデジタル信号に変換処理を行う音響処理装置と、
 音の特徴を学習した音響モデルを記憶した第1の記憶装置と、
 予め冗長語と冗長語以外の通常の単語(通常単語)とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した第2の記憶装置と、

前記音響処理装置及び前記第1及び第2の記憶装置に接続され、前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する装置と、
 を有する音声認識装置。

【請求項29】

アナログ音声の入力装置と、
 前記入力装置に接続され、前記アナログ音声をデジタル信号に変換処理する変換装置と、
 音の特徴を学習した音響モデルを記憶した第1の記憶装置と、
 予め冗長語と冗長語以外の通常の単語(通常単語)とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した第2の記憶装置と、

前記変換装置及び前記第1及び第2の記憶装置に接続され、前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する装置と、

前記認識された結果を表示する表示装置と
 を有するコンピュータ・システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】

本発明は、音声認識装置およびその方法に関するものであり、より具体的には、人の自然な発話を認識して文章化し、冗長語（disfluency）と呼ばれる無意味な単語を自動的に除去してテキストデータを作成する音声認識装置およびその方法に関するものである。

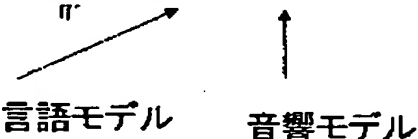
【0002】

【従来の技術】

音響モデルと言語モデルを用いて音声認識を行う統計的方法は知られており、例えば、「A Maximum Likelihood Approach to Continuous Speech Recognition (L.R. Bahl他, IEEE Trans. Vol. PAMI-5, No. 2, 1983, March)」や「単語を認識単位とした日本語の大語彙連続音認識（西村他、情報処理学会論文誌、第40巻、第4号、1999年4月）」がそのような方法について記述している。その概略について説明すると、生成された文章Wが発話され、それが音響処理部において音響処理されて得られた信号からその特徴Xが抽出され、そのX及びWを用いて、以下の式

【数1】

$$W' = \underset{W}{\operatorname{argmax}} P(W|X) = \underset{W}{\operatorname{argmax}} P(W) P(X|W)$$



言語モデル 音響モデル

に従って最適と考えられる認識結果W' が出力され、文章が構成される。つまり、単語列Wが発話されたときの当該特徴量（X）の出現確率P（X | W）とW自身の出現確率（P（W））の積が最大（a r g m a x）となる単語列W' が認識結果として選択される。

【0003】

ここで前者の確率P（X | W）を求めるために音響モデルが用いられ、その確

率の高い単語が認識の候補として選択される。一方、後者の確率 $P(W)$ を近似するためによく用いられるものが言語モデルであり、具体的には N -gram モデルである。これは N 個の連続した単語組の出現確率から文全体、すなわち単語列 W の出現確率を近似する方法であり、次式のように定式化される。

【数 2】

$$\begin{aligned} P(W) &= P(w_0) P(w_1 | w_0) P(w_2 | w_0 w_1) \times \\ &\quad \dots, P(w_n | w_0 w_1, \dots, w_{n-1}) \\ &\equiv P(w_0) P(w_1 | w_0) \prod_{i=2}^n P(w_i | w_{i-2} w_{i-1}) \end{aligned}$$

【0004】

かかる式では、次の単語 $w[n]$ の出現確率が直前の $N-1$ 個の単語にのみ影響を受けるという仮定を行う。 N の値はさまざまなものが考えられるが、その有効性と必要とする学習データのバランスから $N=3$ がよく用いられ、本式も $N=3$ の場合を記述している。以下、 n 個の語からなる文章 W の n 番目の語を $w[n]$ のように表現することとすると、ここでは当該 $N-1$ 個（つまり 2 個）という条件の元での単語 $w[n]$ の出現確率、つまり $P(w[n] | w[n-2] w[n-1])$ の積として単語列 W の出現確率が計算される。ここで、かかる式において、| の左 ($w[n]$) は認識の対象となる単語を示し、| の右 ($w[n-2] w[n-1]$) はその条件となる 2 つ前、1 つ前の単語を示す。さまざまな単語 $w[n]$ についてのそれぞれの出現確率 $P(w[n] | w[n-2] w[n-1])$ はあらかじめ用意されたテキストデータより学習しておき、辞書としてデータベース化して保存しておく。例えば、文の先頭に「単語」という語が出現する確率は 0.0021、その後に「検索」が続く確率は 0.001 等のようにして保存されている。

【0005】

さて、あらかじめ用意された原稿があるような場合については、上述した N -gram モデルで十分であるが、音声認識の適用分野でそのような場合はむしろまれであり、実際はより自然な発話を認識することが応用上重要である。その場合には、内容・意味をもつ通常の単語だけではなく、「アノー」「ソノー」とい

った間投詞的表現や「エー」「アー」といった無意味な単語が発声される。これらは不要語や冗長語と呼ばれるが、それらに対応したN-gramモデルを作成し、それらが自動的に除去されることが望ましい。

【0006】

従来、このために提案されたN-gramモデルの拡張は「透過単語」という概念を導入したものであった。たとえば「単語N-gramモデルを用いた音声認識システムにおける未知語・冗長語の処理（甲斐他、情報処理学会論文誌、第40巻、4号、1999年4月）」や「放送音声の書き起こしに関する検討（西村、伊東、音響学会秋季全国大会、1998年）」にはその方法が記述されている。たとえば前者では「冗長語は文節間において比較的自由に出現するものであり、連接制約であるN-gramが有効に働くとは考えられない」という仮定を元に学習時、認識時いずれにおいても、冗長語の存在を無視して確率計算が行なわれる。たとえば、 $w[n-1]$ を冗長語だとすると $w[n]$ の出現確率は

$$P(w[n] | w[n-2] w[n-1])$$

から計算されるべきところを、 $w[n-1]$ を無視し、 $P(w[n] | w[n-3] w[n-2])$ として推定する。このように無視される、つまりスキップされる単語である冗長語を「透過単語」と呼ぶ。このモデルにおいては、それ自身、つまり冗長語は冗長語以外の単語（通常単語）の各々の間に等しい確率で出現するとして確率の計算が行われる。

【0007】

しかしながら、この冗長語が実際に何ら情報をもっておらず、通常単語間に自由に出現するものか否かについて、英語ではその仮定を否定する報告がある。たとえば「Statistical Language Modeling for Speech Disfluencies (A. Stolcke, E. Shriberg, Proc. of ICASSP96)」では、冗長語に対しても通常のN-gramを適用した結果、冗長語に後続する単語の予測精度が、透過単語モデルより向上したことが記述されている。だが先の透過単語の説明で明らかなように、経験上、冗長語のもつ性質が通常単語と異なり、単純な順序列としてモデル化することが最適ではないことも知られている。

【0008】

一方、現在一般に用いられている、ディクテーション用音声認識システムでは、異なる2つ以上の言語モデルを補間するという方法が用いられることが多い。これは本来ベースとなる汎用の言語モデルでは、コンピュータ、スポーツといった分野固有の文に対して十分対応できない場合、その分野固有のテキストから学習した特定分野言語モデルを汎用のそれと組み合わせる手法であり、以下のように確率計算が行われる。

$$\Pr(w[n] | w[n-2], w[n-1]) = \lambda P1(w[n] | w[n-2], w[n-1]) + (1-\lambda) P2(w[n] | w[n-2], w[n-1])$$

ただし、P1は汎用言語モデルを示し、P2は特定分野言語モデルを示す。ここで λ は補間係数であり、実験により最適な値に設定することとされている。

【0009】

【発明が解決しようとする課題】

本願発明は、従来より認識率の高い音声認識装置及び音声認識方法を提供することを目的とする。

【0010】

さらに、本願発明は、冗長語周辺部における単語予測の精度を向上させることを目的とする。

【0011】

【課題を解決するための手段】

本願発明では、単語を冗長語とそれ以外の通常の単語にわけ、予測される単語、条件となる先行単語いずれにおいても、この2つを区別し、それぞれについて上述の補間法を冗長語を含む単語列に適用することによって冗長語周辺部における単語予測の精度を向上させる。

【0012】

より具体的には、アナログ音声入力信号をデジタル信号に変換処理を行う音響処理手段と、音の特徴を学習した音響モデルを記憶した記憶手段と、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶手段と、前記デ

デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する手段と、を有する音声認識装置を提供する。

【0013】

また、アナログ入力手段と、前記アナログ音声をデジタル信号に変換処理する手段と、音の特徴を学習した音響モデルを記憶した記憶手段と、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶手段と、前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する手段と、前記認識された結果を表示する表示装置とを有するコンピュータ・システムを提供する。

【0014】

さらに、アナログ音声入力信号をデジタル信号に変換処理し、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶し、前記デジタル信号について音響モデルと前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する音声認識方法を提供する。

【0015】

さらに、アナログ音声を入力し、前記アナログ音声をデジタル信号に変換処理し、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶し、前記デジタル信号について音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識し、前記認識された結果を表示する音声認識方法を提供する。

【0016】

さらに、コンピュータ・プログラムを有するコンピュータ読みとり可能な記憶

媒体であって、前記記憶媒体は、音響モデルと、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶し、さらに前記コンピュータ・プログラムは、コンピュータに入力されたアナログ音声入力信号が変換されたデジタル信号について前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させるものである記憶媒体を提供する。

【0017】

さらに、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶媒体を提供する。

【0018】

さらに、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章について音声認識をするための装置であって、認識の対象となる入力された語が通常単語かどうか判断する手段と、通常単語であると判断された場合、さらに入力された語を認識するために必要な条件となる語が通常単語のみからなるかどうかを判断する手段と、条件となる語が通常単語のみでなく冗長語も含むと判断された場合、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させる手段と、を有する音声認識装置を提供する。

【0019】

さらに、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章について音声認識をするための方法であって、認識の対象となる入力された語が通常単語かどうか判断するステップと、通常単語であると判断された場合、さらに入力された語を認識するために必要な条件となる語が通常単語のみからなるかどうかを判断するステップと、条件となる語が通常単語のみでなく冗長語も含むと判断さ

れた場合、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させるステップと、を有する音声認識方法を提供する。

【0020】

さらに、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章について音声認識をするためのコンピュータ・プログラムを有するコンピュータ読みとり可能な記憶媒体であって、前記コンピュータ・プログラムは、コンピュータに認識の対象となる入力された語が通常単語かどうか判断する手順と、通常単語であると判断された場合、さらに入力された語を認識するために必要な条件となる語が通常単語のみからなるかどうかを判断する手順と、条件となる語が通常単語のみでなく冗長語も含むと判断された場合、冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識させる手順と、を実行させるものである記憶媒体を提供する。

【0021】

さらに、アナログ音声入力信号をデジタル信号に変換処理を行う音響処理装置と、音の特徴を学習した音響モデルを記憶した第1の記憶装置と、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した第2の記憶装置と、前記音響処理装置及び前記第1及び第2の記憶装置に接続され、前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する装置と、を有する音声認識装置を提供する。

【0022】

さらに、アナログ音声の入力装置と、前記入力装置に接続され、前記アナログ音声をデジタル信号に変換処理する変換装置と、音の特徴を学習した音響モデル

を記憶した第1の記憶装置と、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した第2の記憶装置と、前記変換装置及び前記第1及び第2の記憶装置に接続され、前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する装置と、前記認識された結果を表示する表示装置とを有するコンピュータ・システムを提供する。

【0023】

【発明の実施の形態】

図1は、本願発明の構成を示すブロック図である。ブロック101において生成された文章（すなわち真の文章）Wは、Sとして発話される（102）。一般にこの文章生成及び発話は発話者によってなされる。発話されたSは、文章Wだけでなく、「アノー」「エート」等の冗長語を含む音である。この入力された音声であるSは、音声認識手段110中の音響処理部111によって信号Xに変換されて記憶される。この変換された信号Xは、言語復号部112において、音の特徴を学習した音響モデル113と、後述する学習により予め作成された言語モデルについての辞書114を用いて、真の文章だけでなく冗長語を含むものから必要な言語を抽出して認識結果W'とする。このような一連の動作を音にエンコードされたデータから必要なデータをデコード（復号）するともいう。そして、認識された結果の文章を表示する（120）。

【0024】

次に、本願発明の実施される装置の典型的なシステムの一例について図2に示す。発話者が発話した音声はマイク210からアナログ信号として入力され、コンピュータ装置220のサウンドカード221によりデジタル信号に変換処理され、メモリ222に記憶される。なお、このような機能を奏するものであれば、ハードウェアとして実現されていても、ソフトウェアとして実現されていてもいずれでも良い。さらに、音響モデル及び後述する学習により作成された言語モデルを含む辞書もメモリに記憶されている。これらの変換され記憶された信号及び

辞書からCPU 223は言語の復号を行い、認識された結果を表示装置230に表示する。なお、マイクはコンピュータ装置や表示装置と一体化されていることもある。また、表示装置による結果の表示はCRTや液晶などディスプレイだけでなく、印刷装置により紙に印刷されるような場合もある。

【0025】

図1のブロック図と図2のシステムとの対応を考えると、一例として、マイク、サウンドカード及びメモリにより音響処理部の前段部が実現され、また音響モデルと辞書（言語モデル）はメモリ上に記憶され、音響処理部の後段部と言語復号はCPUを用いて行われ（音響処理部の処理は単なるデジタル信号への変換だけでなく、特徴Xの抽出などのためにCPUが必要となる）、文章の表示は表示装置において行われる。ただし、これらの関係は固定的なものではない。例えば変換された信号を記憶するメモリと辞書を有するメモリは物理的に異なるものとすることもできる。ただし、このような場合もそれらを一体としてメモリとして考えることも可能である。

【0026】

上述の通り、本発明は単語を冗長語とそれ以外の通常の単語（通常単語）にわけ、予測される単語、条件となる先行単語いずれにおいても、この2つを区別し、さらにこの補間法を冗長語を含む単語列に適用することによって、冗長語周辺部における単語予測の精度を向上させようとするものであるが、その具体的な内容について以下により詳細に述べる。

【0027】

冗長語のもつ性質の特殊性は、その情報が後続単語の予測に寄与する程度が不明確な点にある。すなわち、冗長語を無視して、それ以前の通常単語から予測する方がよいというのが従来技術で述べた透過単語モデルの考え方である。これに対して、通常のN-gramモデルは、後続単語の予測に最も役にたつのは、それに隣接する単語であるという考え方に基づいている。そこで、この2つの方法で複数の言語モデルを作成し、それらのモデルを補間して予測を行う。具体的には以下の過程をへて確率の学習・計算を行う。以下簡単のためすべてN=3（3-gram）とする。

【0028】

まず、予め準備されたテキストデータに基づいて、学習により図3に示すような辞書300を生成する。具体的には、以下のようにして2つのモデルに基づく辞書が生成される。なお、いずれの場合においても、冗長語が連続する場合には、たとえば連続する冗長語列を1つの冗長語に置き換えて学習すればよい。

【0029】

1. 各通常単語について、冗長語を除去した通常単語のみからなる学習テキストデータを用いて3-gramの確率を学習する。つまり上述の透過単語モデルの考え方による言語モデルである。これをモデルU(310)とする。

【0030】

2. 冗長語を含むテキストについて学習を行う。この場合、さらに以下の2つの場合がある。ここでwfilは冗長語を表す。

(1) 予測の対象となる語が冗長語wfilである場合の3-gram確率、すなわち $P(wfil|w[n-2], w[n-1])$ を学習する。ここで、例えば条件のうち $w[n-1]$ が通常単語、 $w[n-2]$ が冗長語というような可能性があるが、その場合には $w[n-2]$ をとばし、 $w[n-3]$ を条件部に用い $P(wfil|w[n-1], w[n-3])$ とする。つまり、この場合条件に冗長語は含まないようにする。

(2) 予測の対象となる語が通常単語であり、その直前が冗長語である場合、その冗長語のみを条件とする2-gram確率 $P(w[n]|wfil)$ を学習する。つまり、通常単語の直前が冗長語である場合の確率である。

以上のように、条件部、予測対象のそれぞれにおいて冗長語と通常単語を分けた確率を学習することが本学習の基本である。この2. (1)、(2)をあわせてモデルD(320)とする。

【0031】

このように学習により生成された辞書を用いて、図4に示すようなフローチャートに従って認識のための確率計算を行う。以下、かかる図4について説明する。

【0032】

まず、音響処理部によって変換された音声信号について音響モデルを用いた計

算の結果に基づいて認識候補として単語を選択する(400)。ここで、例えば認識候補としての単語は数百個程度に絞られる。次に、候補となる単語が通常単語であるか、冗長語であるかが判断される(410)。本発明では、対象単語が冗長語であるか通常単語であるかによって異なる確率計算を行うからである。

【0033】

候補の単語が通常単語 $w[n]$ である場合、条件部 $w[n-2]w[n-1]$ も通常単語のみからなるかどうか判断され(420)、条件部 $w[n-2]w[n-1]$ も通常単語のみからなる場合は、モデルUの $P(w[n] | w[n-2], w[n-1])$ から $w[n]$ の予測、つまり確率計算を行う(430)。

【0034】

認識対象の単語が通常単語 $w[n]$ であるが、条件部に冗長語が存在すると判断された場合は、モデルUとモデルDの両方の辞書を用いて確率計算が行われる(440)。

【0035】

このブロック440について、本実施例についてより詳細に説明すれば、認識対象の単語が通常単語 $w[n]$ であるが、条件部に冗長語が存在するかどうか判断され(510)、直前の単語 $w[n-1]$ が冗長語の場合は、それをとばして最も近接した通常単語まで遡りそれを条件としたモデルUの確率と、直前が冗長語であったという条件のモデルDの確率を補間して $w[n]$ の確率計算を行う(520)。つまり当該確率 Pr は以下の式により求められる。

$$Pr = \lambda PU(w[n] | w[n-2], w[n-1]) + (1 - \lambda) PD(w[n] | wfil)$$

ここで、 PU ：モデルUによる確率、 PD ：モデルDによる確率を示す。また、補間係数 λ は予め実験により最適な値に決定する。例えば、補間係数の決定は、実験において補間係数 λ を0から1まで0.1刻みで変化させ、冗長語を含むモデル文章について最も認識率が高くなるような値を採用する。

【0036】

認識対象の単語が冗長語 $w[n]$ である場合、条件部 $w[n-2], w[n-1]$ に冗長語が存在するにはそれをとばして(530)、最も近接した通常単語 $w[n-i], w[n-j]$ まで遡り、それを条件部とするモデルD、 $PD(w[n] | w[n-i], w[n-j])$ により確率計算

を行う（540）。

【0037】

以上の結果に基づいて、言語モデルによる予測される語の確率を求める（460）。この場合、最も確率の高い語を認識語として表示させるようにしても良いし、あるいはこの予測結果からさらに候補語を百語程度に絞り込み、再度音響モデルによる確率計算を詳細に行い認識結果を算出しても良い。

【0038】

このように、本発明においては、確率計算において、予測対象、条件のいずれにおいても、冗長語である場合と、そうではない場合を分離し、それらを別々に学習しておくこと、そして確率計算時においては、その2つのモデルを補間して通常単語・冗長語の影響がより適した割合で考慮されることになる。

【0039】

なお、本発明においては、補間係数 λ を適切な値に設定することが、その効果に大きく影響する。その値に影響を与える要素としては、不要語の出現率、学習コーパスの量など多くのものがあり、理論的考察が困難なため、実験により、言語モデルがどの程度有効に働いているかという値を求め、それを基準として決定することが多い。

【0040】

言語モデルの有効性は一般にパープレキシティという統計量が用いられる。その定義の詳細はたとえば、「音声・音情報のディジタル信号処理」（鹿野他著、昭晃堂、1997年）に記述があるが、直感的には全認識対象語彙（たとえば6万語）が言語モデルを使用することにより、何単語程度の対象語彙と等価にまで減少したかを意味し、より小さい値の方がよい。

【0041】

たとえばテレビの講演を書き起こした約100万単語からなるコーパスの場合、 λ は0.2が最適であった。そこでパープレキシティを求めた結果、透過単語モデルは225.0であるのに対して、本発明によるモデルでは195.1であり、約13%の改善であることがわかった。当業者であれば理解されるところであるが、この分野における改良案には不要語に限らず多くのものがあるが、その中でこの値は大

きいものである。

【0042】

以上の例の他、冗長語予測においても条件部において、直前単語が冗長語、すなわち冗長語が連続する場合に、その条件付き確率と補間することなど、さまざまな組み合わせが考えられる。

【0043】

例えば、上述の例では、モデルDについて、対象語の直前が冗長語の場合についてのみ学習を行わせたが、例えばその対象語の2語前が冗長語の場合（3-gramモデル）等の確率もさらに加味して補間する方法もある。

【0044】

さらに、本発明においては前述のように、音響処理部によって変換された音声信号について、まず先に音響モデルを用いた計算の結果に基づいて認識候補として単語を選択してある程度絞り込んだ上でそれらに言語モデルを適用して確率を求めて総合判定を行い認識の最終結果を得るが（図6（a））、逆にまず先に認識された語の履歴に基づいて言語モデルを適用して候補となる語をある程度絞り込んだ上でそれらに音響モデルを適用して認識を行う方法も考えられる（図6（b））。

【0045】

【発明の効果】

本願発明により、従来より認識率の高い音声認識装置及び音声認識方法を提供するが可能となり、さらに、冗長語周辺部における単語予測の精度を向上させることが可能となる。

【図面の簡単な説明】

【図1】

本発明の構成を示すブロック図である。

【図2】

本発明の実施されるコンピュータ・システムの一例を示す図である。

【図3】

本発明に用いられる辞書についての図である。

【図 4】

本発明の処理の流れを示すフローチャートである。

【図 5】

本発明の処理の流れを示すフローチャートである。

【図 6】

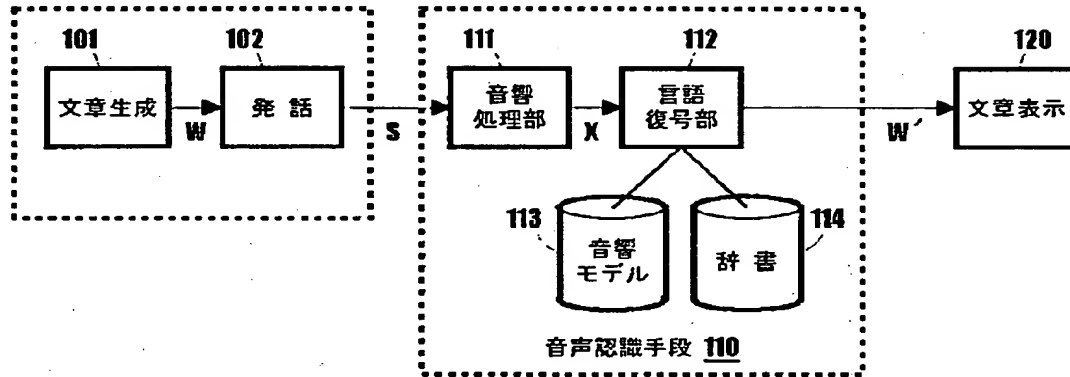
音声認識処理の流れを示すフローチャートである。

【符号の説明】

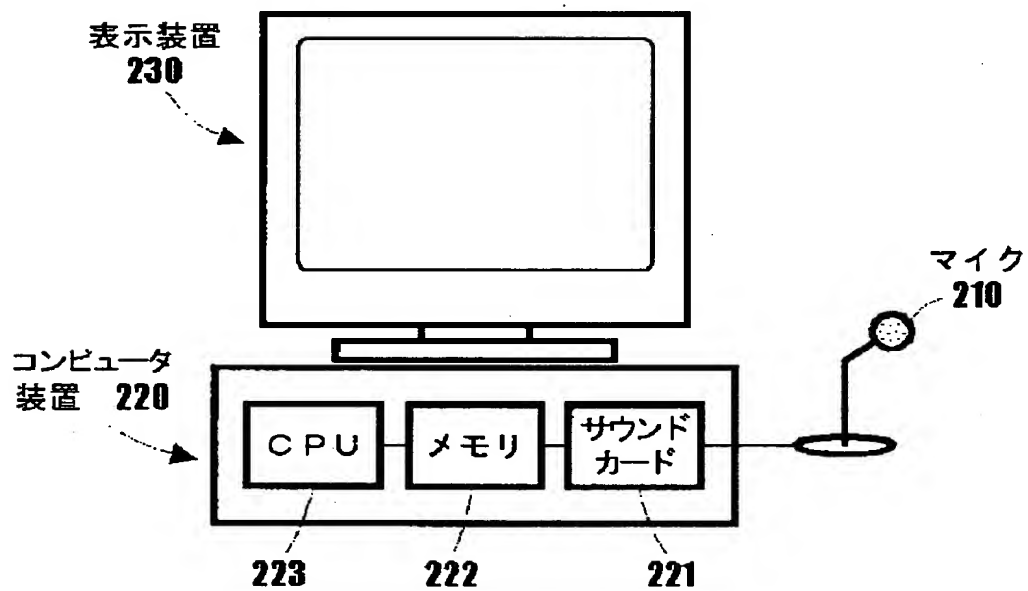
- 1 1 0 音声認識手段
- 2 1 0 マイク
- 2 2 0 コンピュータ装置
- 2 2 1 サウンドカード
- 2 2 2 メモリ
- 2 2 3 CPU
- 2 3 0 表示装置

【書類名】 図面

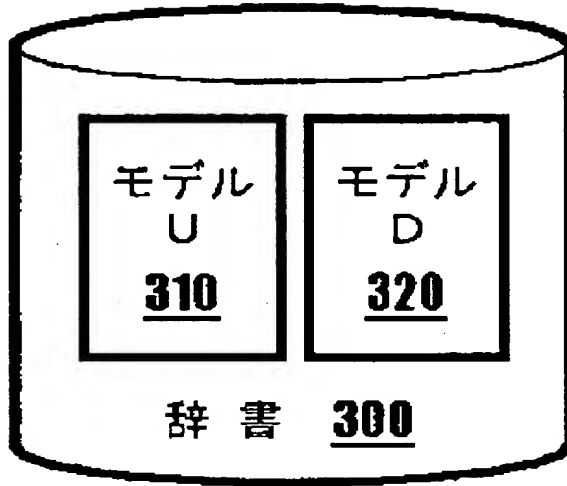
【図 1】



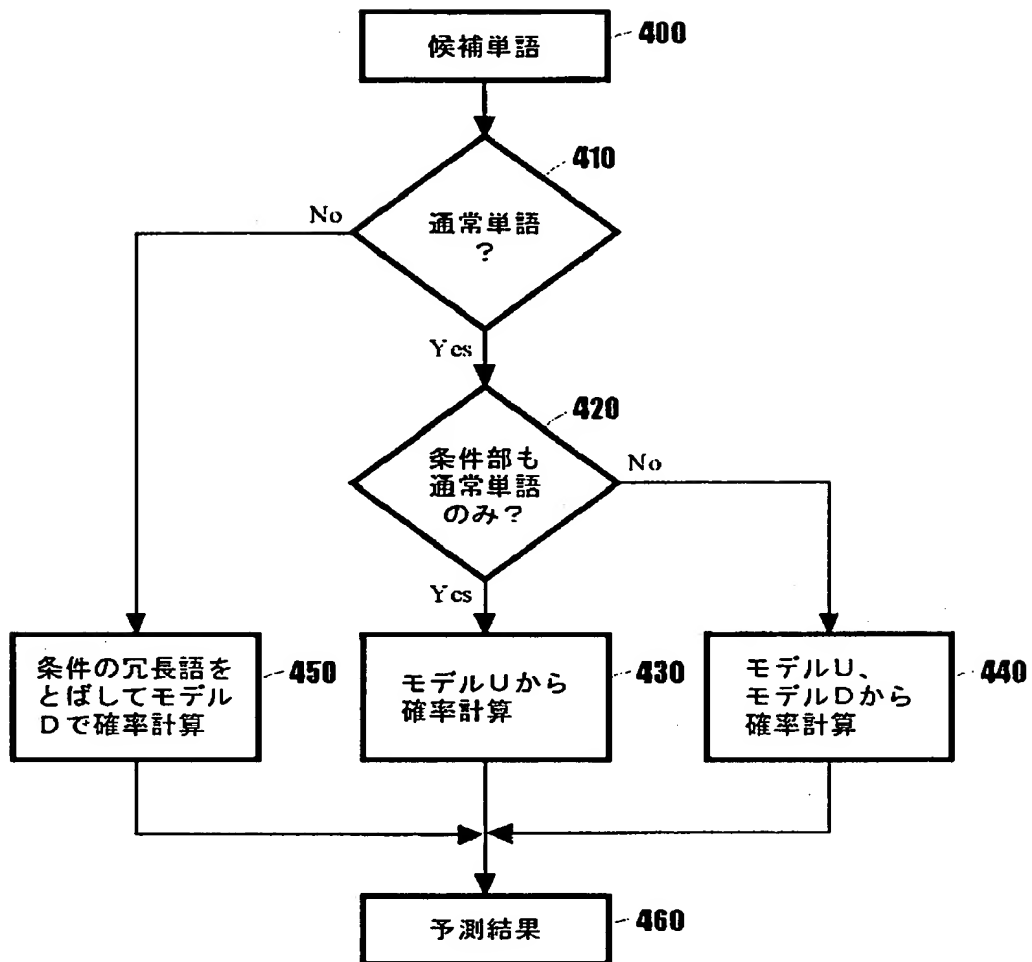
【図 2】



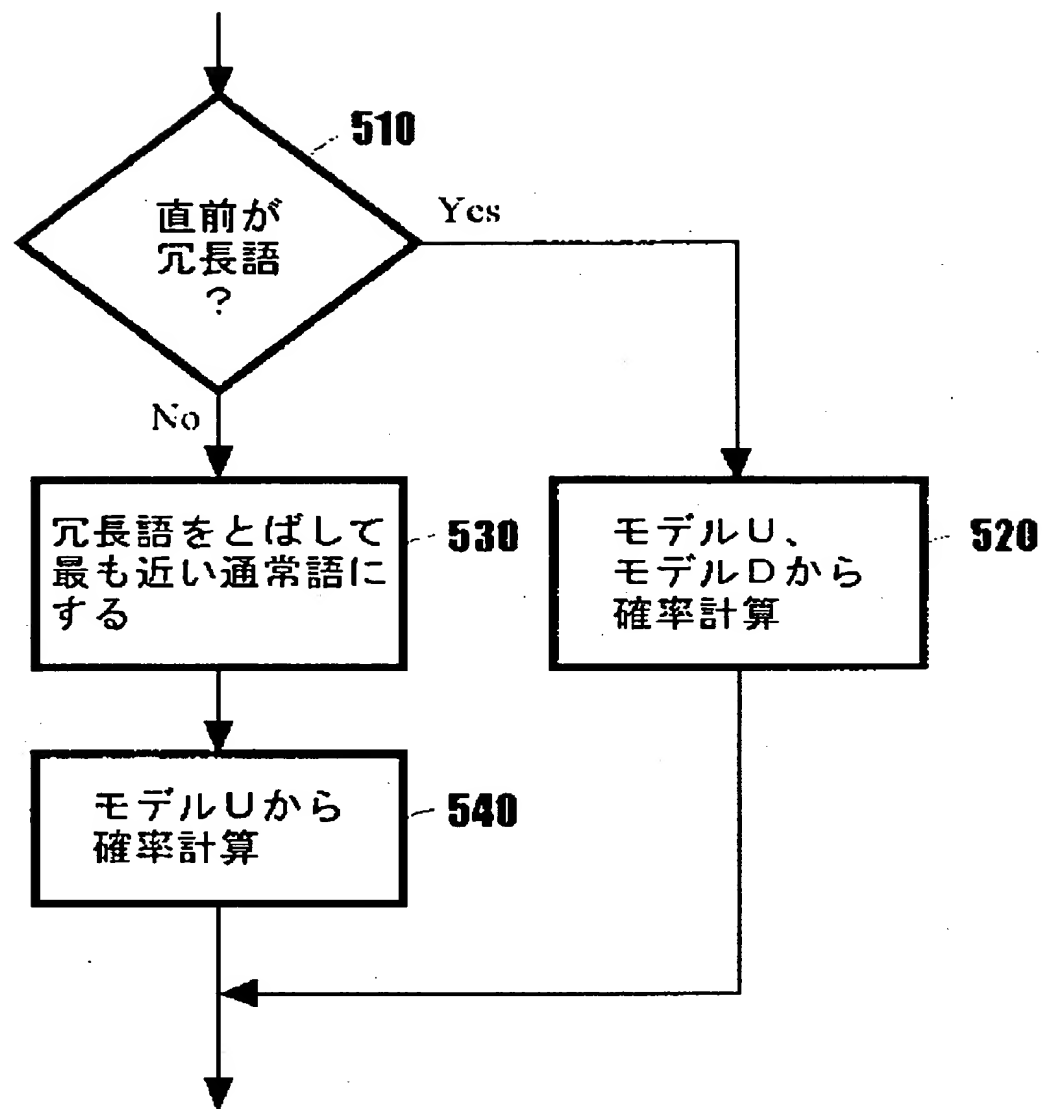
【図 3】



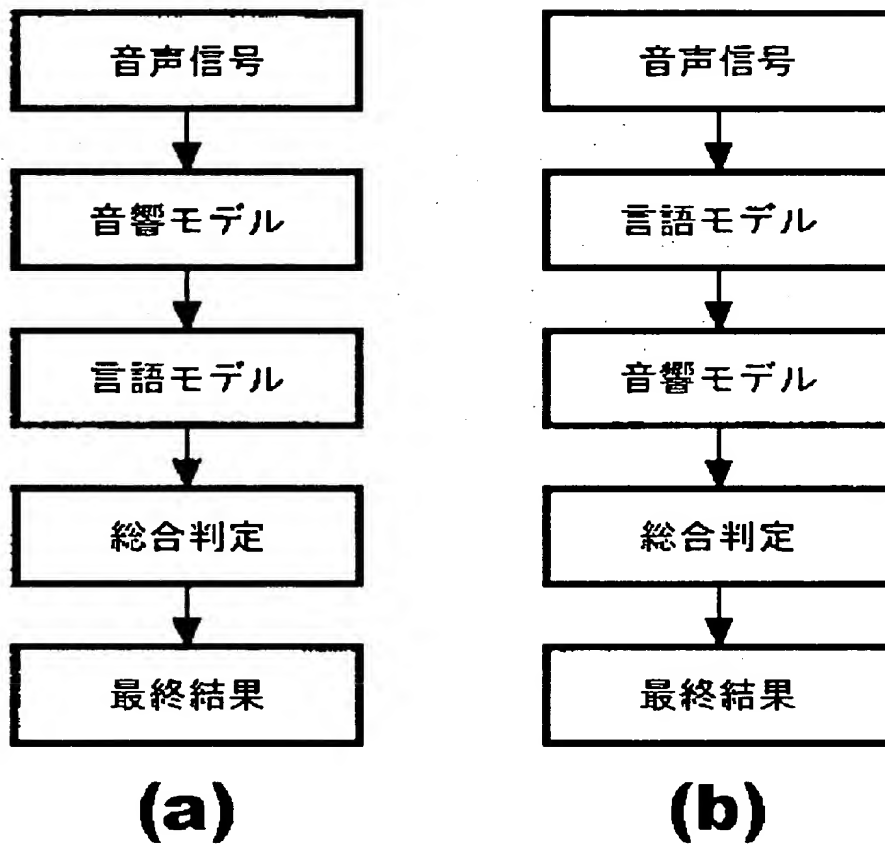
【図 4】



【図 5】



【図 6】



【書類名】要約書

【要約】

【課題】

従来より認識率の高い音声認識装置及び方法を提供すること。

【解決手段】

本願発明では、単語を冗長語とそれ以外の通常の単語にわけ、予測される単語、条件となる先行単語いずれにおいても、この2つを区別して予測を行うことにより冗長語周辺部における単語予測の精度を向上させる。そのために、アナログ音声入力信号をデジタル信号に変換処理を行う音響処理手段と、音の特徴を学習した音響モデルを記憶した記憶手段と、予め冗長語と冗長語以外の通常の単語（通常単語）とを含む文章に基づいて学習した第1の言語モデルと、冗長語を無視して通常単語のみの文章に基づいて学習した第2の言語モデルと、の両方の言語モデルを有する辞書を記憶した記憶手段と、前記デジタル信号について前記音響モデル及び前記辞書を用いて確率を計算して最も確率の高い語を入力された音声として認識する手段と、を有する音声認識装置を提供する。

【選択図】図1

認定・付加情報

特許出願の番号	平成11年 特許願 第370413号
受付番号	59901273268
書類名	特許願
担当官	第七担当上席 0096
作成日	平成12年 1月 4日

<認定情報・付加情報>

【提出日】	平成11年12月27日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [390009531]

1. 変更年月日 1990年10月24日
[変更理由] 新規登録
住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレイション